Faculty Working Papers

A MONTE CARLO COMPARISON OF TRADITIONAL AND STEIN-RULE ESTIMATORS UNDER SQUARED ERROR LOSS

Thomas A. Yancey, George Judge, and
Mary E. Bock

#211

College of Commerce and Business Administration
University of Illinois at Urbana-Champaign

FACULTY WORKING PAPERS

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

October 3, 1974

A MONTE CARLO COMPARISON OF TRADITIONAL AND STEIN-RULE ESTIMATORS UNDER SQUARED ERROR LOSS

Thomas A. Yancey, George Judge, and Mary E. Bock

#211

A Monte Carlo Comparison of Traditional and Stein-Rule Estimators under Squared Error Loss

Thomas A. Yancey, George Judge, and Mary E. Bock*

University of Illinois at Urbana-Champaign

1. Introduction

Almost two decades ago Stein [9] demonstrated for the problem of estimating the K dimensional parameter vector $\underline{\beta}$ for the linear statistical model $\underline{y} = X\underline{\beta} + \underline{e}$, that under a squared error loss measure of goodness it is possible to improve on the maximum likelihood estimator (MLE), $\underline{b}=X'\underline{y}$, if the dimension of the unknown coefficient vector is greater than two. As is traditional in the Stein form of the statistical model $\underline{y}$ and $\underline{e}$ are (Tx1) normal vectors with means $X\underline{\beta}$ and $\underline{o}$, respectively, and covariance matrix $\sigma^2 I_T$ and X is a (TxK) design matrix which either by choice or canonical reduction is such that $X'X = I_K$. At the beginning of the 1960's, for this same problem, James and Stein [5] specified a Stein rule estimator (SRE), $\underline{\tilde{\beta}} = (1-c^*/u)(\underline{b}-\underline{\beta}_o) + \underline{\beta}_o = (1-(K-2)(T-K)K^{-1}(T-K+2)^{-1}/u)(\underline{b}-\underline{\beta}_o)+\underline{\beta}_o$, which dominated the MLE.[1/] In the SRE $u = (\underline{b}-\underline{\beta}_o)'(\underline{b}-\underline{\beta}_o)/K\hat{\sigma}^2$ is the traditional test statistic, distributed as F with (T-K) and K degrees of freedom, and $\underline{\beta}_o$ is a K dimensional hypothesis vector. Shortly thereafter, Baranchik [1] and Stein [8] proposed the Stein positive rule estimator (SPRE), $\underline{\tilde{\beta}}^* = I_{[tc^*,\infty)}(u)\ (1-tc^*/u)(\underline{b}-\underline{\beta}_o) + \underline{\beta}_o$, for $o \le t \le 2$, that was uniformly superior to the SRE, $\underline{\tilde{\beta}}$. As Efron and Morris [4, pp. 123, 124] have shown for the SPRE, the rules with $t < 1$ are dominated and the rules $1 \le t \le 2$ do not dominate one another. In a recent paper Sclove, Morris and Radhakrishnan [7] demonstrated that the traditional pre test estimator (PTE),

---

*The authors wish to acknowledge the help of William Harris and Robert Peterson in carrying out the computations reported in this paper.

1/ Using a squared error loss measure for evaluating estimator performance an estimator $\underline{\hat{\theta}}$ is said to be superior to or to dominate another estimator $\underline{\tilde{\theta}}$, if for all $\underline{\theta}$, the risk function for $\underline{\hat{\theta}}$ is equal to or less than that of $\underline{\tilde{\theta}}$, i.e., $E(\underline{\hat{\theta}}-\underline{\theta})'(\underline{\hat{\theta}}-\underline{\theta}) - E(\underline{\tilde{\theta}}-\underline{\theta})'(\underline{\tilde{\theta}}-\underline{\theta}) \le \underline{o}$, with strict inequality holding for some $\underline{\theta}$.

$\hat{\underline{\beta}} = I_{(o,c)}(u)\underline{\beta}_o + I_{[c,\infty)}(u)\underline{b} = I_{[c,\infty)}(u)(\underline{b}-\underline{\beta}_o) + \underline{\beta}_o$, where $c = F^{\alpha}_{(K,T-K)}$, is uniformly inferior to a Stein modified positive rule estimator (SMPRE), $\tilde{\tilde{\underline{\beta}}} = I_{[c,\infty)}(u)I_{[tc^*,\infty)}(u)(1-tc^*/u)(\underline{b}-\underline{\beta}_o) + \beta_o$. The SMPRE $\tilde{\tilde{\underline{\beta}}}$, is in reality a preliminary test estimator involving the K dimensional hypothesis vector $\underline{\beta}_o$ and the SPRE, $\tilde{\underline{\beta}}^*$, and has the form, $\tilde{\tilde{\underline{\beta}}} = \underline{\beta}_o$, if $u < c$ and, $\tilde{\tilde{\beta}} = \tilde{\underline{\beta}}^*$, if $u \geq c$.

In spite of the apparent superiority of the Stein rules there has been no rush by applied workers to abandon maximum likelihood procedures, and in fact, in economics one is hard put to find applications of these estimation rules in the econometric literature. Possibly one reason for the reluctance to change estimators may be uncertainty relative to the magnitude of the risk gains of changing estimation rules. For example, in the Sclove, et al. paper [7] the risk comparisons for the SMPR estimator are not evaluated since they involve complicated hypergeometric functions. With the hope of removing some of the uncertainty relative to the performance of variants of the Stein-rule, the modest purpose of this paper is to report the results of using Monte Carlo sampling experiments to characterize the nature of the risk functions for the SPR and SMPR estimators and study their performance relative to that of the conventional and pre test estimators.

2. Analytical Risk Comparisons

As is well known the maximum likelihood estimator has risk, when $X'X = I$

(2.1) $$E[(\underline{b}-\underline{\beta})'(\underline{b}-\underline{\beta})] = \rho(\underline{b},\underline{\beta}) = \sigma^2 K.$$

The risk for SRE, $\tilde{\underline{\beta}}$, may be expressed as

(2.2) $$\rho(\tilde{\underline{\beta}},\underline{\beta}) = \sigma^2 K - (K-2)^2(T-K)/(T-K+2)\; E[1/(K-2+2H)]$$

where H is a Poisson distributed random variable with mean $\lambda = (\underline{\beta}-\underline{\beta}_o)'(\underline{\beta}-\underline{\beta}_o)/2\sigma^2$. The difference in the risk functions of the MLE and SRE may be expressed analytically as

$$(2.3)\quad \rho(\underline{b},\underline{\beta}) - \rho(\underline{\tilde{\beta}},\underline{\beta}) = (K-2)^2(T-K)/(T-K+2)\ E\ [1/K-2+H]$$

$$=(K-2)^2(T-K)/(T-K+2)\ e^{-\lambda/2}\sum_{i=o}^{\infty}(\lambda/2)^i/(K+2+2i)i!$$

$$=(K-2)^2(T-K)/(T-K+2)e^{-\lambda/2}[1/K-2 + \lambda/2K + \lambda^2/8(K+2) + \ldots]$$

$$\geq(K-2)^2(T-K)/(T-K+2)e^{-\lambda/2}\sum_{i=o}^{N}(\lambda/2)^i/(K+2+2i)i!,$$

where N is an integer $\geq$ o.

The risk for the pre test estimator, $\underline{\hat{\beta}}$, as derived by Sclove, et. al. [7] and Bock, et al. [3], may be expressed as

$$(2.4)\quad E[(\underline{\hat{\beta}}-\underline{\beta})'(\underline{\hat{\beta}}-\underline{\beta})] = \sigma^2K - \sigma^2Kp_1 - (p_2-2p_1)(\underline{\beta}-\underline{\beta}_o)'(\underline{\beta}-\underline{\beta}_o)$$

where p_1 and p_2 are probabilities that the ratios of independent random variables, $X^2_{(K+2+2H)}/X^2_{(T-K)}$ and $X^2_{(K+4+2H)}/X^2_{(T-K)}$ are equal to or less than $cK/(T-K)$, where H has a Poisson distribution with mean, λ. The $\rho(\underline{\hat{\beta}},\underline{\beta}) \leq \rho(\underline{b},\underline{\beta})$ if $\lambda = (\underline{\beta}-\underline{\beta}_o)'(\underline{\beta}-\underline{\beta}_o)/2\sigma^2 \leq K/4$.

The risk for the SPR estimator, $\underline{\tilde{\beta}}^*$, from the work of Baranchik [1] and Stein [8], may be expressed as

$$(2.5)\quad E[(\underline{\tilde{\beta}}^*-\underline{\beta})'(\underline{\tilde{\beta}}^*-\underline{\beta})] = \rho(\underline{\tilde{\beta}},\underline{\beta}) - 2(\underline{\beta}-\underline{\beta}_o)'(\underline{\beta}-\underline{\beta}_o)\ E[I_{(o,c)}(p_3)(c^*/p_3)]$$

$$- \sigma^2\ E[I_{(o,c)}(p_4)(1 - c^*/p_4)^2\ X^2_{(K+2H)}$$

where p_3 and p_4 are the ratios, $(T-K)\ X^2_{(K+2H+2)}/K\ X^2_{(T-K)}$ and $(T-K)X^2_{(K+2H)}/K\ X^2_{(T-K)}$, respectively, of independent Chi square variables and H is Poisson random variables with mean, λ.

The pre test estimator inadmissibility proof of Sclove, et. al. [7] means that when we compare the risks of the PTE, $\underline{\hat{\beta}}$, and the SMPRE, $\underline{\tilde{\tilde{\beta}}}$, we must evaluate the function

$$(2.6)\quad o \leq (E[(\hat{\underline{\beta}}-\underline{\beta})'(\hat{\underline{\beta}}-\underline{\beta})] - E[(\tilde{\tilde{\underline{\beta}}}-\underline{\beta})'(\tilde{\tilde{\underline{\beta}}}-\underline{\beta})])/\sigma^2$$

$$= E[(tc^*K/(T-K))\, X^2_{(T-K)}\{ 1/X^2_{(K+2,\lambda)}\; I_{(c,\infty)}\; (((T-K)/K)\, X^2_{(K+2,\lambda)}$$

$$/X^2_{(T-K)})(-(t\, c^*K)/(T-K)(X^2_{(T-K)}\,/X^2_{(K+2,\lambda)}) + 2K + 4\lambda)$$

$$+(2\lambda/X^2_{(K+4,\lambda)})I_{[c,\infty)}(((T-K)/K)\, X^2_{(K+4,\lambda)}/X^2_{(T-K)})$$

$$(-((tc^*K)/(T-K))(X^2_{(T-K)}/X^2_{(K+4,\lambda)}) +2\}],$$

where $X^2_{(\ell,\lambda)}$ denotes a Chi square random variable with ℓ degrees of freedom and non centrality parameter λ.

Since this function is difficult to evaluate precisely, Monte Carlo sampling experiments were performed to create empirical risk functions for the estimators and to determine the differences in the risk functions for alternative critical values of the test, c, and tc*. If the SRE, $\tilde{\underline{\beta}}$, is substituted for the SPRE, $\tilde{\underline{\beta}}$, in the SMPRE, $\tilde{\tilde{\underline{\beta}}}$, that is $\tilde{\tilde{\underline{\beta}}}_1 = I_{(c,\infty)}(u)(1-c^*/u)(\underline{b}-\underline{\beta}_o) + \underline{\beta}_o$, the difference between the two PT estimators can, in lieu of (2.3), be evaluated and in fact sets a lower limit for the difference in risk between the PTE and SMPRE.

3. The Sampling Experiment

In order to generate information concerning the nature of the risk function for the SMPRE, data were generated using the following orthonormal statistical model

$$(3.1)\quad \underline{y} = 1349.4\underline{x}_o + 49.72\underline{x}_1 + 81.43\underline{x}_2 + 38.27\underline{x}_3 + 7.62\underline{x}_4 + \underline{e},$$

where the e's are normally and independently distributed with mean zero and variance 60.8 and risk, $E[(\underline{b}-\underline{\beta})'(\underline{b}-\underline{\beta})] = 304.00$ and $X'X = I$.

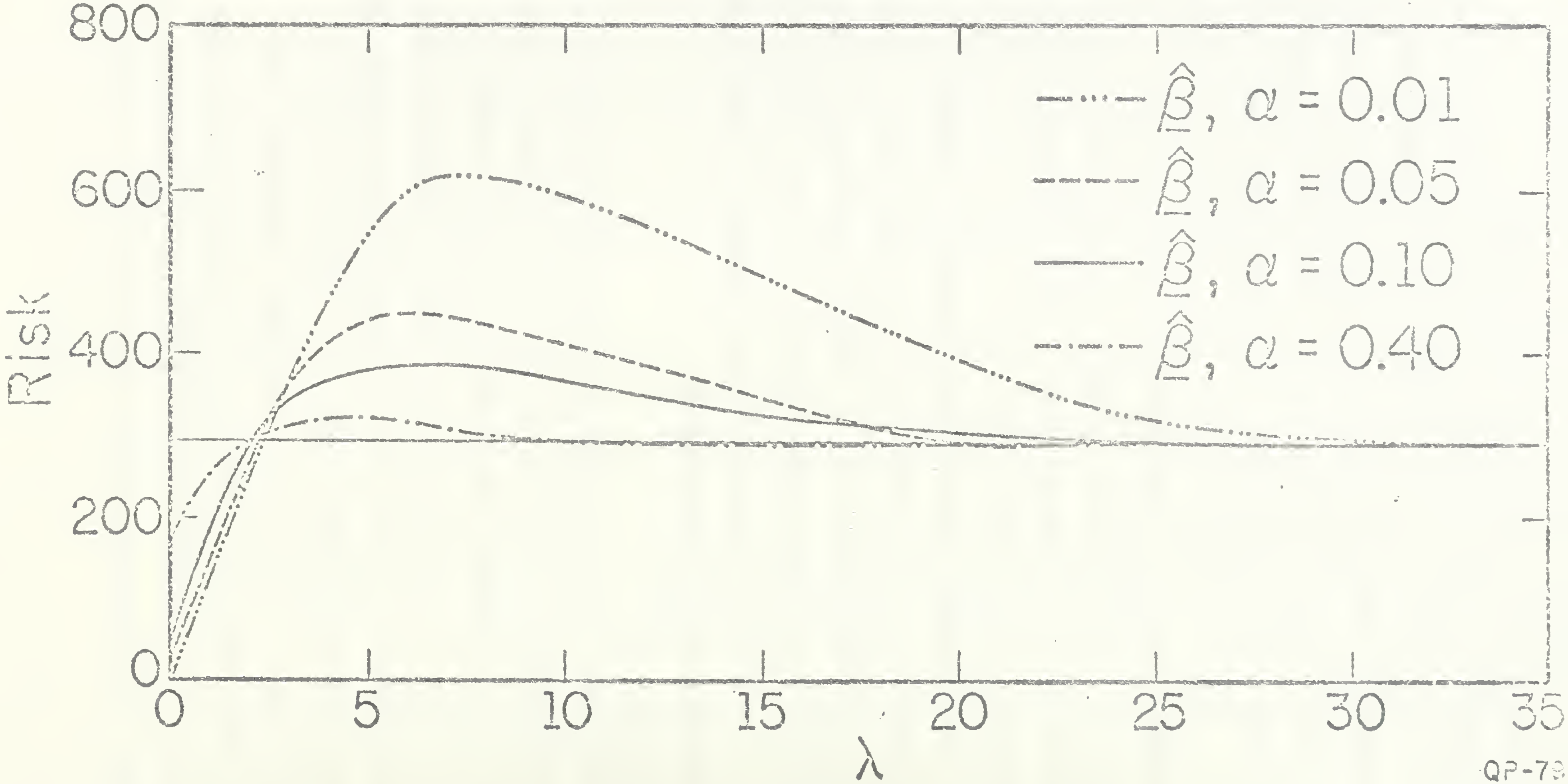
Three sets of one hundred and twenty samples of size 10, 15, and 25 were generated using a single X design matrix, for each sample size, in conjunc-

tion with e's which were based on Wold's table of random normal variables. The e's used in the experiment are the same as those reported by Neiswanger and Yancey [6] and have a mean of .0237 and a variance of 61.89. The Kolmogorov-Smirnov test statistic, under the hypothesis that the e's were generated from a normal distribution, N(0. 60.8), was well below the five percent level of statistical significance. Indeed, means differing by .0237 or more could be expected about 87 percent of the time.

Least squares estimates for the 120 samples of size 10, 15, and 25 were obtained and the corresponding values of the test statistic were computed for a range of incorrectly specified hypotheses. The degree of hypothesis misspecification was tabulated in terms of the noncentrality parameter for the F distribution, $\lambda = (\underline{\beta}-\underline{\beta}_o)'(\underline{\beta}-\underline{\beta}_o)/2\sigma^2$, and this parameter was specified to take on discrete values over a range of λ from 0 to 45. Critical values for typical levels of the test (α = .1, .05, .01) and a non typical level of the test (α = .4) were used for the conventional and Stein rule pre test estimators. Since it is not possible to determine the optimal (minimum risk) choice of the critical value tc* for $1 \leq t \leq 2$, in the Stein-rule positive part estimator, the risk functions for the SMPRE were developed for $tc^* = c^* = (T-K)\, K^{-1}\, (K-2)\, (T-K+2)^{-1}$ and other tc*'s with $c^* \leq tc^* \leq 2c^*$. Only the results for samples of size 25 are graphically represented.

4. The Sampling Results

The empirical risk functions of the least squares (ML) and the pre test (PT) estimators are given in Figure 1 for samples of size 25. Significance levels for the test statistic of .01, .05, .1, and .4 were used and the results are plotted in Figure 1. To provide one measure of validity for the experiment, the analytic and empirical risk functions were compared for both the ML and PT estimators. In this experiment, the value of the empirical least squares risk function of 293.3 is 96.5 per cent of that of the analytic risk function which



QP-78

Figure 1. Empirical Risk Functions for the Maximum Likelihood and Pre Test Estimators.

is 304.0. The empirical preliminary test risk is nearly always below the analytic PT risk function and for both the .01 and .05 levels of significance is almost always from 95 to 98 per cent of the analytic risk function throughout the range of λ that was explored. Furthermore, the empirical pre test and least squares risk functions do intersect in the interval from 1.25 to 2.5 for λ as analytical results indicate they should.

Since the main focus of this paper is directed to comparisons of the empirical risk functions of the James-Stein (SR) Stein positive rule (SPR), and modified Stein-rule (SMPR) estimators with each other as well as with those of least squares (ML) and the pre test (PT) estimators, it is to these results that we now turn.

a. The Empirical SRE Risk Functions.

A comparison of the SR and ML estimators is shown in Figure 2 for samples of size 25. The optimum value of $tc^* = (T-K)\ (K-2)/(T-K+2)K$ was used for the (SR) estimator and as expected from analytical results (e.g. Efron & Morris, [4]) the largest gains in the SRE relative to the MLE occur for the smaller values of the non-centrality parameter λ. The risk functions of the SRE varied from 41 per cent of that of the MLE, when $\lambda= 0$, to 98 per cent of MLE risk when $\lambda=35$. The SRE risk function was .53, .90, and .95 per cent of the ML risk function for values of λ of 1, 5, and 15, respectively, and was 98.3 per cent of ML risk for $\lambda=45$.

When the sample size is reduced to 10 the SR risk function varied from 50 per cent of ML risk for $\lambda = 0$ to 99 per cent of ML risk when $\lambda=35$. The upward shift of the SR risk relative to ML risk as T decreases is predicted analytically by the SR risk function (Stein[8, p. 352]). The upward shift of the SR risk function, as T decreases from 25 to 10 observations, is a modest fraction of a percentage point for, $\lambda=1$, but for, $\lambda<.5$, the gains ranged up to 9 percentage points.

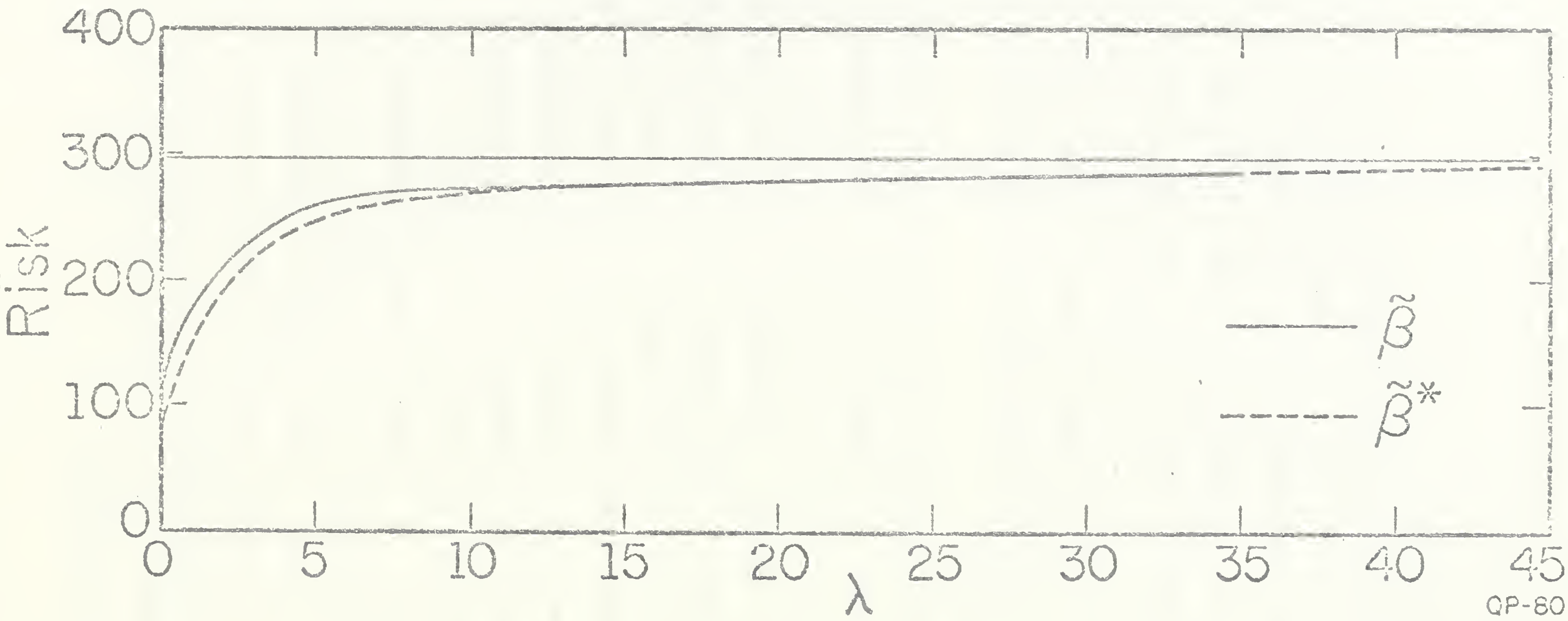
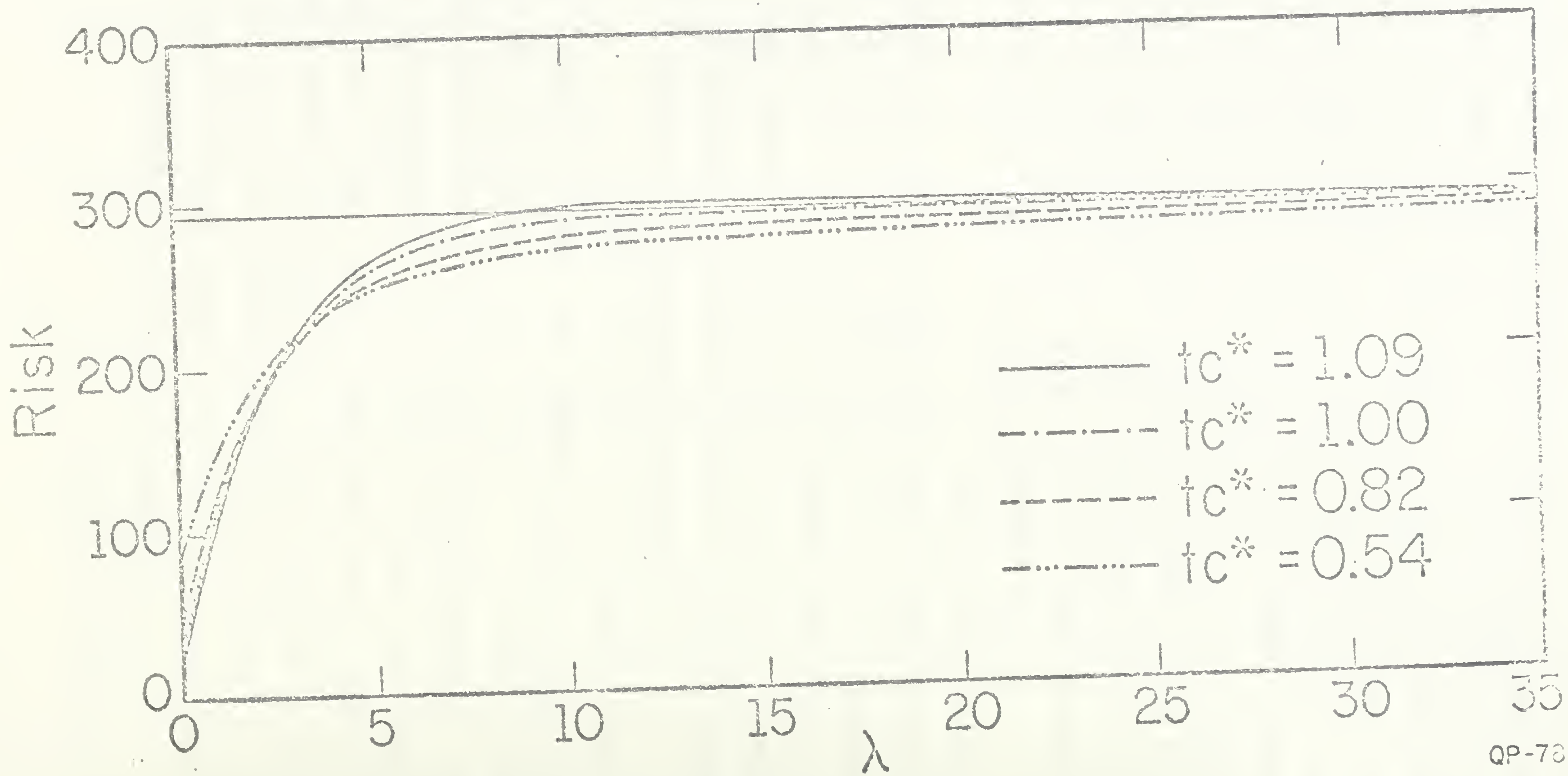


Figure 2. Empirical Risk Functions for the Maximum Likelihood, James and Stein and Stein Positive Rule Estimators

b. The SPRE Risk Functions.

Although as Baranchik [7] and Stein [8] have shown, the positive part estimator dominates the James-Stein estimator, there is very little information as to whether or not the reduction in risk the SPRE produces is non trivial in nature. In this experiment the improvement of the risk functions of the SPRE and the SRE over MLE was substantial for the smaller values of λ. For samples of size 25, the SPRE and SRE risk functions are given in Figure 2. The case plotted used the optimum tc^* for the SRE (i.e. $tc^* = (K-2)(T-K)/(T-K+2)K$) for both estimators). In line with analytical results, the effects of other values of tc^* from c^* to $2c^*$ (i) shifted the risk functions for the SRE upward (James and Stein [5] and (ii) showed as expected that tc^* for $t \varepsilon [1,2]$ the SPRE risk functions do not dominate one another and for different tc^* with $t \varepsilon [0,1]$, the SPRE risk functions are dominated (Efron and Morris [4], pp. 123-24]). For expository purposes the SPRE risk functions for four values of tc^*, for $t \varepsilon [1,2]$, are given in Figure 3. As would be expected both the SR and the SPR estimators lost their risk advantage over the MLE except for small values of λ, ($\lambda \leq 9$ for SPRE and $\lambda \leq .25$ for SRE) when tc^* was $2(K-2)(T-K)/(T-K+2)K$.

In comparing the risk functions of the SPRE and SRE for samples of size 25 and $tc^* = c^*$, the SPRE risk function varied from 73 per cent of the SRE risk function, for $\lambda=0$, to being identical with the SRE risk for $\lambda>10$. The ratio of SPRE risk to SRE risk was .83 at the value $\lambda=1$ and about .95 at the value $\lambda=5$. A comparison of the SPRE risk function to that of the MLE for samples of size 25, indicates that the ratio of risk is .30 for $\lambda=0$, .52 for $\lambda=1$, .85 for $\lambda=5$, and .93 for $\lambda=10$. When the sample size was reduced to $T=10$, the ratio of SPRE risk to SRE risk decreased to .83 at $\lambda=0$ with the ratio converging to 1 for $\lambda=10$.



QP-78

Figure 3. Empirical Risk Functions for the Stein Positive Rule Estimator for $c^* = (K-2)(T-K)K^{-1}(T-K+2)^{-1}$ and $1 \leq t \leq 2$.

c. The SMPRE Risk Functions.

In a 1972 paper Sclove et al., i) modified the Stein plus rule estimator to permit the c in the indicator function to differ from $c^* \leq tc^* \leq 2(T-K)(K-2)/(T-K+2)K$ in the remainder of the estimator and ii) proved that for comparable values of α the SMPR estimator would dominate the PT estimator. Written in the form of Section 1, the c in the indicator function takes values associated with the levels of significance used in the PT estimator. The significance levels for the PTE are usually for small values of α such as .01, .05, .10. The α values associated with the tc*'s for which the SMPRE is a minimax estimator, i.e., sominates the MLE, for samples of size 25, ranged from .4 to .74 in this experiment.

The risk functions of the SMPRE as well as those of the PTE, for samples of size 25, α=.01 and .4, and $tc^* = (T-K)(K-2)/(T-K+2)K$, are shown in Figure 4. For the case α=.01 the ratio of the empirical SMPRE risk to that of the PTE risk rises sharply from .81 for $\lambda=0$, to a high of .98 when $\lambda=2$ then declines to .94 when $\lambda=10$ and rises back to .98 at $\lambda=35$. For the situation α=.40, the ratio of the SMRE and PTE risks rises steadily from .45 at $\lambda=0$ to .98 at $\lambda=35$.

In this experiment, the relative advantage of the SMPRE over the PTE in terms of risks increased as α increased from .01 to .74. These gains are small for all levels for $\lambda \geq 15$. When the sample size was reduced to 10 and α=.01 the risk functions for both the SMPRE and PTE are virtually identical in value for all λ. The risk functions started near zero for $\lambda=0$, and rose sharply to values from 5 to 5.5 times that for the MLE risk and were still four times the MLE risk for $\lambda=50$. Even with α=.10 both the SMPRE AND PTE risk functions began close to 0 for, $\lambda=0$, rose sharply reaching maximum values about twice those for the MLE, risk at $\lambda=9$ and declined to about the MLE risk level for $\lambda=50$. The risk function of SMPRE remained below that of the PTE in all cases as analytical results indicate it should. As

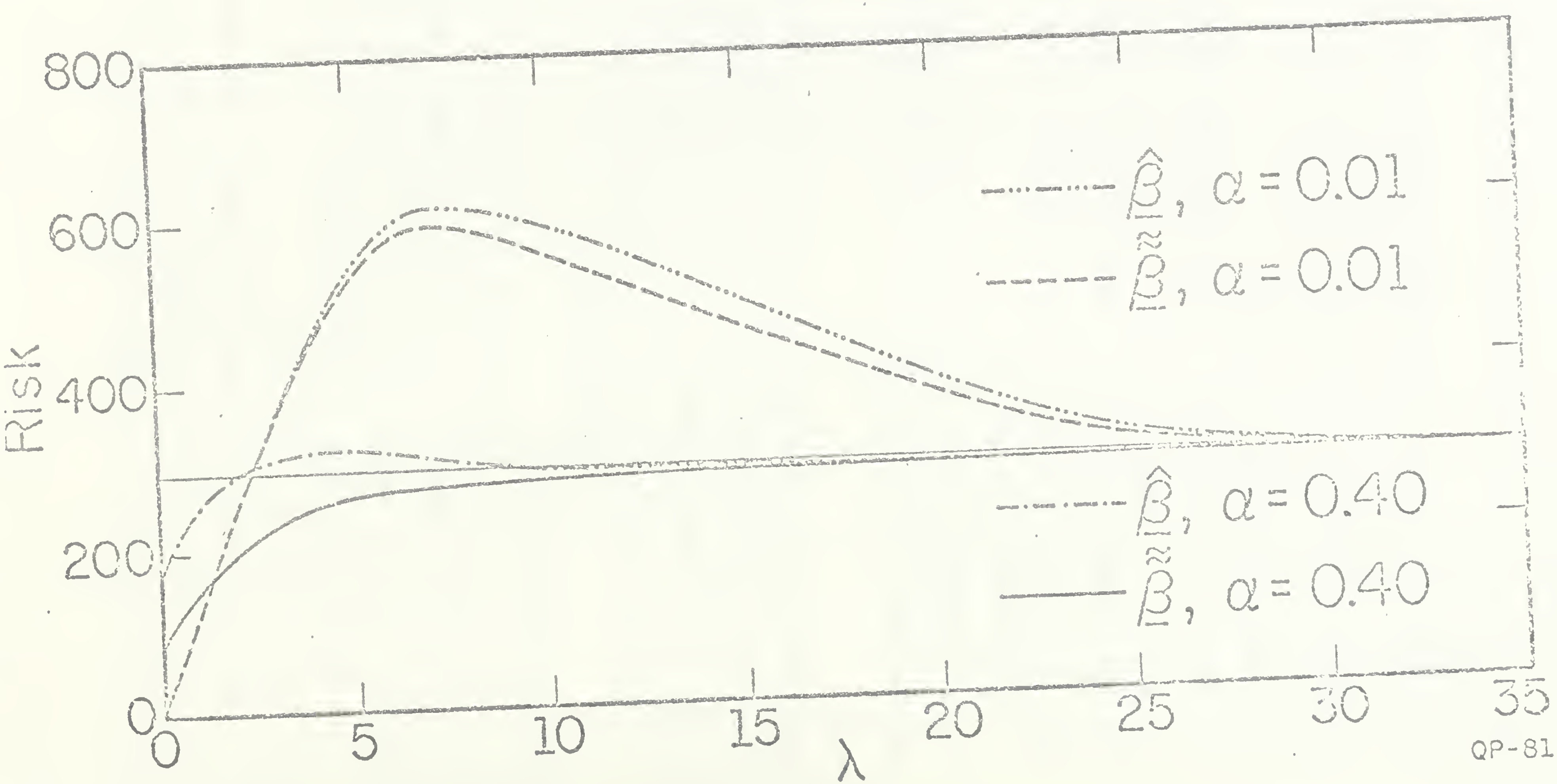


Figure 4. Empirical Risk Functions for the Pre Test and Sclove Modified Stein Rule Estimators Under Two Levels of Statistical Significance

expected, the risk functions of the SMPRE for samples of size 10 and $tc^* = (T-K)(K-2)/(T-K+2)K$, approached that of the SPRE as c decreased and α increased. With $\alpha \geq .535$ the empirical SMPRE risk function remained below that of the MLE for values of λ through 50.

For samples of both size 10 and 25, decreasing c from the critical value associated with $\alpha = .01$, while holding tc^* at c^*, altered the shape of SMPRE risk function by increasing the risks for $\lambda \leq .50$ reducing risks for $\lambda \geq 2$. Most of the effect on the shape and level of the SMPRE risk function of changing α, when $\alpha \geq .50$, is for $\lambda \leq 1$.

5. Concluding Remarks

Although care must be taken in generalizing the results of a limited sampling study, our experiment, comparing various Stein-rule estimators with each other and with least squares and the preliminary test estimator under squared error loss for the orthonormal regression model case suggests the following:

i) The choice of the level of the test has a dramatic impact on the nature of the risk function for the modified Stein-rule and preliminary test estimators.

ii) For each level of the test the risk functions for the modified Stein-rule and the pre test estimators have the same general shape over the range of λ explored; the difference between the values taken by the two risk functions are greatest for small values of λ and "large" values of α.

iii) The problem of the optimal value of tc^* for $t \in [1,2]$ and the optimal level of the test c in the Sclove modified positive rule estimator remains, as in the case of the pre test estimator, to be resolved.

iv) For levels of α in excess of those normally employed in practice, making use of a variant of the Stein-rule family appears

to lead to a significantly lower risk function over much of the empirically useful part of the parameter space.

v) The risk gains of the SMPRE relative to the PTE under traditional levels of significance appear to be small, if not trivial.

As Bock [2] has shown, the analytic results for the orthornormal case do extend to the general case if the trace of the inverse of the X'X matrix of the explanatory variables divided by its largest characteristic root is greater than 2. The risk behavior of the Stein-rule estimators for the non orthonormal case when this condition does not hold remains to be determined.

Strawderman [10] has developed minimax admissible rules. Although we know these rules do not dominate the Stein positive rule estimator, little is known relative to how they compare over the range of the parameter space to the Stein rule variants discussed in this paper.

6. References

[1] Baranchik, A. J. (1964): "Multiple Regression and the Estimation of the Mean of a Multivariate Normal Distribution," Stanford University Technical Report, No. 51.

[2] Bock, M. E. (1973): "Minimax Estimators of the Mean of a Multivariate Distribution," forthcoming, Annals of Statistics.

[3] Bock, M. E., T. A. Yancey, and G. G. Judge (1972): "The Statistical Consequences of Preliminary Test Estimators in Regression," Journal of American Statistical Association, 68:109-116.

[4] Efron, B. and C. Morris (1973): "Stein's Estimation Rule and Its Competitors--An Empirical Bayes Approach," Journal of American Statistical Association, 68:117-130.

[5] James, W. and C. Stein (1961): "Estimation with Quadratic Loss," Proceedings of Fourth Berkeley Symposium on Mathematical Statistic Problems, University of California Press, Berkeley, pp. 361-379.

[6] Neiswanger, W. A. and T. A. Yancey (1959): "Parameter Estimates and Autonomous Growth," Journal of American Statistical Association, 54:389-402.

[7] Sclove, S. [illegible], and R. Radhakrishnan (1972): "Non-Optimality of Preliminary-Test Estimators for the Multinormal Mean," Annals of Mathematical Statistics, 43:1481-90.

[8] Stein, C. ([illegible]): "An Approach to the Recovery of Interblock Information in Balanced [illegible] Block Designs," Research Papers in Statistics: Festschrift [illegible], F. N. David (Ed.), John Wiley and Sons, Inc., New York, [illegible].

[9] Stein, C. (19[illegible]): "Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution," Proceedings of the Third Berkeley Symposium, University of California Press, Berkeley, 1:197-206.

[10] Strawderman, W. E. (1971): "Proper Bayes Minimax Estimators of the Multivariate Normal Mean," Annals of Mathematical Statistics, 42:385-88.